

特開平11-143346

(43)公開日 平成11年(1999)5月28日

(51) Int. Cl. ⁶	識別記号	府内整理番号	F I	技術表示箇所
G09B 5/04			G09B 5/04	
G10L 3/00			G10L 3/00	A
	551			551 E
9/00	301		9/00	301 A

審査請求 未請求 請求項の数11 OL (全13頁)

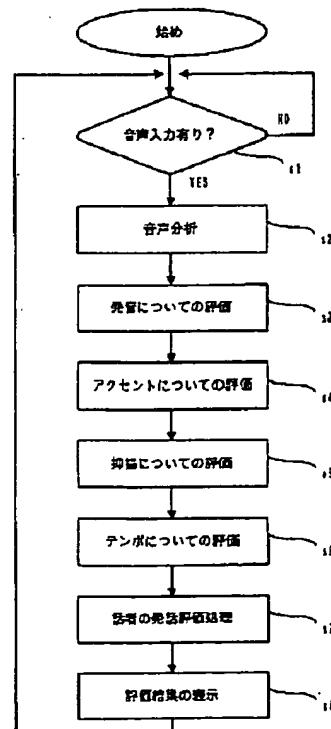
(21)出願番号	特願平9-303202	(71)出願人	000002369 セイコーエプソン株式会社 東京都新宿区西新宿2丁目4番1号
(22)出願日	平成9年(1997)11月5日	(72)発明者	宮沢 康永 長野県諏訪市大和3丁目3番5号 セイコ ーエプソン株式会社内
		(74)代理人	弁理士 鈴木 喜三郎 (外2名)

(54)【発明の名称】語学練習用発話評価方法およびその装置並びに発話評価処理プログラムを記憶した記憶媒体

(57)【要約】

【課題】 英会話など他国の言語による会話を練習する場合、特徴的な要素を比較することによって客観的な評価を行い効率の良い練習を可能とする。

【解決手段】 手本の持っている各発話内容ごとに、その発話内容について各フレームごとに音声分析し、複数の要素としての発音、アクセント、抑揚、テンポ、を表すデータを予め求めておく。そして、話者の発話した発話内容について各フレームごとに音声分析し（ステップs 2）、前記複数の要素に対するデータを求め、この話者のデータと前記手本のデータとを、パターンマッチングによる各フレーム間の対応付けを行い、発音についての評価（ステップs 3）、アクセントについての評価（ステップs 4）、抑揚についての評価（ステップs 5）、テンポについての評価（ステップs 6）を行ったのち話者の発話に対する評価値を出力する（ステップs 7, s 8）。



【特許請求の範囲】

【請求項1】 手本の発話内容に従って発話した話者の音声を分析して得られたデータと前記手本の発話内容に対するデータとを比較して、話者の発話評価を行う語学練習用発話評価方法において、

手本の持っているそれぞれの発話内容ごとに、その発話内容についてそれぞれのフレームごとに音声分析し、発話の適正さを判断するに必要な要素に対するデータを求めておき、

話者の発話した発話内容についてそれぞれのフレームごとに音声分析し、前記発話の適正さを判断するに必要な要素に対するデータを求め、この話者のデータと前記手本のデータとを、パターンマッチングによる各フレーム間の対応付けを行ってそれぞれのデータ間を比較し、その比較結果に基づいて話者の発話評価を行うことを特徴とする語学練習用発話評価方法。

【請求項2】 前記発話の適正さを判断するに必要な複数の要素は、発音、アクセント、抑揚、発話のテンポであつて、これらの各要素のうち少なくとも1つを用い、発音についての評価は、前記データとして話者の音声の特徴を表す音声特徴データと手本の音声の特徴を表す音声特徴データを用い、両者の音声特徴データをパターンマッチングによるフレーム間の対応付けを行って比較し、その比較結果を基に話者の発音についての評価を行い、

アクセントについての評価は、前記データとして話者のパワー情報を手本のパワー情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のパワー情報を対応づけて比較し、その比較結果を基にアクセントについての評価を行い、抑揚についての評価は、前記データとして話者のピッチ情報を手本のピッチ情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のピッチ情報を対応づけて比較し、その比較結果を基に抑揚についての評価を行い、

テンポについての評価は、手本の発話内容を幾つかに区切ることによって複数の区間を設定し、それぞれの区間における時間情報を求めておき、話者の発話内容に対しても、前記パターンマッチングによるフレーム間の対応付けを用いて、前記手本の区切りに対応する区切りごとに区切って、それにより設定された区間ごとの時間情報を求め、前記手本における各区間ごとの時間情報と話者の対応する区間における時間情報を比較し、その比較結果をもとに発話のテンポの評価を行うことを特徴とする請求項1記載の語学練習用発話評価方法。

【請求項3】 前記アクセントについての評価を行う際、話者のパワーと手本のパワーとをそれぞれのフレームごとに差を取り、その差分を基にアクセントについての評価を行う処理と、話者のパワーの極大値と手本のパワーの極大値とを比較し、その時間的位置とそのパワー

の大きさを基にアクセントについての評価を行う処理の少なくとも1つの処理を行うことを特徴とする請求項2記載の語学練習用発話評価方法。

【請求項4】 前記話者の発話についての評価を行う際、手本の発話内容を幾つかに区切って複数の区間を設定し、話者の発話内容に対しても、前記パターンマッチングによるフレーム間の対応付けを用いて、前記手本の区切りに対応する区切りごとに区切って、それにより設定された区間ごとに、前記発話の適正さを判断するに必要な要素についての評価を行うとともに、全区間に渡って前記発話の適正さを判断するに必要な要素についての評価を行い、さらに、その発話内容全体に対する総合的な評価を行うことを特徴とする請求項1から3のいずれかに記載の語学練習用発話評価方法。

【請求項5】 前記話者の発話評価を行う際、話者の音声と手本の音声に対して話者正規化処理を施したのち発話評価処理を行うことを特徴とする請求項1から4のいずれかに記載の語学練習用発話評価方法。

【請求項6】 手本の発話内容に従って発話した話者の音声を分析して得られた音声データと前記手本の発話内容に対する音声データとを比較して、話者の発話評価を行う語学練習用発話評価装置において、手本の持っているそれぞれの発話内容についてそれぞれのフレームごとに求められた発話の適正さを判断するに必要な要素に対するデータを、それぞれの発話内容ごとに記憶する手本音声データ記憶部と、話者の音声を入力する音声入力部と、この音声入力部に入力された話者の発話した発話内容について、それぞれのフレームごとに音声分析し、前記発話の適正さを判断するに必要な要素に対するデータをフレームごとに求める入力音声分析部と、この入力音声分析部により求められた話者のデータと前記手本のデータとを、パターンマッチングによるそれぞれのフレーム間の対応付けを行ってデータ間を比較し、その比較結果に基づいて話者の発話についての評価を行う発話評価処理部と、この発話評価処理部による評価内容に基づいた評価結果を出力する評価結果出力部と、を有することを特徴とする語学練習用発話評価装置。

【請求項7】 前記発話の適正さを判断するに必要な要素は、発音、アクセント、抑揚、発話のテンポであつて、これらの各要素のうち少なくとも1つを用い、前記発話評価処理部が行う発音についての評価は、前記データとして話者の音声の特徴を表す音声特徴データと手本の音声の特徴を表す音声特徴データを用い、両者の音声特徴データをパターンマッチングによるフレーム間の対応付けを行って比較し、その比較結果を基に話者の発音についての評価を行い、アクセントについての評価は、前記データとして話者のパワー情報を手本のパワー情報を用い、前記音声特徴デ

ータの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のパワー情報を対応づけて比較し、その比較結果を基にアクセントについての評価を行い、抑揚についての評価は、前記データとして話者のピッチ情報と手本のピッチ情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のピッチ情報を対応づけて比較し、その比較結果を基に抑揚についての評価を行い、

テンポについての評価は、手本の発話内容を幾つかに区切ることによって設定された複数の区間ごとの時間情報と、前記パターンマッチングによるフレーム間の対応付けにより前記手本の区切りに対応して区切られて設定された話者の発話内容に対するそれぞれの区間ごとの時間情報を比較し、その比較結果をもとに発話のテンポの評価を行うことを特徴とする請求項6記載の語学練習用発話評価装置。

【請求項8】 前記発話評価処理部は、アクセントについての評価を行う際、話者のパワーと手本のパワーとをそれぞれのフレームごとに差を取り、その差分を基にアクセントについての評価を行う処理と、話者のパワーの極大値と手本のパワーの極大値とを比較し、その時間的位置とそのパワーの大きさを基にアクセントについての評価を行う処理の少なくとも1つの処理を行うことを特徴とする請求項7記載の語学練習用発話評価装置。

【請求項9】 前記発話評価処理部は、手本の発話内容を幾つかに区切って複数の区間を設定し、話者の発話内容に対して、前記パターンマッチングによるフレーム間の対応付けを用いて、前記手本の区切りに対応する区切りごとに区切って、それにより設定された区間ごとに、前記発話の適正さを判断するに必要な要素についての評価を行うとともに、全区間に渡って前記発話の適正さを判断するに必要な要素についての評価を行い、さらに、その発話内容全体に対する総合的な評価を行うことを特徴とする請求項6から8のいずれかに記載の語学練習用発話評価装置。

【請求項10】 前記発話評価処理部の前段に、話者の音声と手本の音声に対して話者正規化処理を施す話者正規化処理手段を設け、この話者正規化処理手段により話者正規化処理を行ったのち、発話評価処理を行うことを特徴とする請求項6から9のいずれかに記載の語学練習用発話評価装置。

【請求項11】 手本の発話内容に従って発話した話者の音声を分析して得られたデータと前記手本の発話内容に対するデータとを比較して、話者の発話評価を行う発話評価処理プログラムを記憶した記憶媒体において、前記発話評価処理プログラムは、

手本の持っているそれぞれの発話内容ごとに、その発話内容についてそれぞれのフレームごとに音声分析し、発話の適正さを判断するに必要な要素に対するデータを求めて記憶させておき、

話者の発話した発話内容についてそれぞれのフレームごとに音声分析し、前記発話の適正さを判断するに必要な要素に対するデータを求め、この話者の一タと前記手本のデータとを、パターンマッチングによる各フレーム間の対応付けを行ってデータ間を比較し、その比較結果に基づいて話者の発話評価を行うことを特徴とする発話評価処理プログラムを記憶した記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は英会話などを練習する場合、ユーザの発話内容に対して評価を行う語学練習用発話評価方法およびその装置並びに発話評価処理プログラムを記憶した記憶媒体に関する。

【0002】

【従来の技術】 最近、英語など他国言語での会話を勉強する人が増えてきている。従来、他国言語の会話を練習する場合、専門の学校に通ったりするのが一般的であるが、自分で勉強する人も多い。

【0003】

しかし、他国言語の会話は、発音やアクセント、抑揚、発話のテンポなどが日本語とは大きく異なるため、一人での練習ではなかなか上達しないのが現状である。そこで、従来より、一人で練習するための会話練習機やそのソフトなども幾つか開発されてきている。たとえば、カセットテープなどで手本となる先生の発話する内容を、ユーザが真似て発話するものもその一つである。

【0004】

【発明が解決しようとする課題】 しかしながら、従来のこの種の会話練習機やそのソフトは、手本の発話内容を自分なりに真似て発話することにより、それぞれの言語独特の発音、アクセント、抑揚、会話のテンポを覚えようとするものが殆どであり、ユーザの発話した会話について客観的な評価を与えてくれるものでは少ない。さらに、それぞれの言語に特有の発音、アクセント、抑揚、テンポなどそれぞれについてきめ細かく評価するというものはなかった。したがって、ユーザにとっては、自分の発話に対する発音、アクセント、抑揚、テンポなどが本当に適正なものかを判断できないのが現状であった。

【0005】

そこで本発明は、ユーザの発話した会話内容に対して、発音、アクセント、抑揚、テンポなど、発話の適正さを判断するための指標となる複数の要素について客観的な評価を行い、且つこれらを総合した評価を与えることを可能とし、効率的な他国言語の会話練習を可能とする語学練習用発話評価方法およびその装置並びに発話評価処理プログラムを記憶した記憶媒体を提供することを目的とする。

【0006】

【課題を解決するための手段】 前述したような目的を達成するために、本発明の請求項1に記載の語学練習用発話評価方法は、手本の発話内容に従って発話した話者の

音声を分析して得られたデータと前記手本の発話内容に対するデータとを比較して、話者の発話評価を行う語学練習用発話評価方法において、手本の持っているそれぞれの発話内容ごとに、その発話内容についてそれぞれのフレームごとに音声分析し、発話の適正さを判断するに必要な要素に対するデータを予め求めておき、話者の発話した発話内容についてそれぞれのフレームごとに音声分析し、前記発話の適正さを判断するに必要な要素に対するデータを求め、この話者のデータと前記手本のデータとを、パターンマッチングによる各フレーム間の対応付けを行ってそれぞれのデータ間を比較し、その比較結果に基づいて話者の発話評価を行うことを特徴としている。

【0007】そして、前記発話の適正さを判断するに必要な複数の要素は、発音、アクセント、抑揚、発話のテンポであって、これらの各要素のうち少なくとも1つを用い、発音についての評価は、前記データとして話者の音声の特徴を表す音声特徴データと手本の音声の特徴を表す音声特徴データを用い、両者の音声特徴データをパターンマッチングによるフレーム間の対応付けを行って比較し、その比較結果を基に話者の発音についての評価を行い、アクセントについての評価は、前記データとして話者のパワー情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のパワー情報を対応づけて比較し、その比較結果を基にアクセントについての評価を行い、抑揚についての評価は、前記データとして話者のピッチ情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のピッチ情報を対応づけて比較し、その比較結果を基に抑揚についての評価を行い、テンポについての評価は、手本の発話内容を幾つかに区切ることによって複数の区間を設定し、それぞれの区間ににおける時間情報を求めておき、話者の発話内容に対しても、前記パターンマッチングによるフレーム間の対応付けを用いて、前記手本の区切りに対応する区切りごとに区切って、それにより設定された区間ごとの時間情報を求め、前記手本における各区間ごとの時間情報と話者の対応する区間における時間情報を比較し、その比較結果をもとに発話のテンポの評価を行うようにしている。

【0008】さらに、前記アクセントについての評価を行う際、話者のパワーと手本のパワーとをそれぞれのフレームごとに差を取り、その差分を基にアクセントについての評価を行う処理と、話者のパワーの極大値と手本のパワーの極大値とを比較し、その時間的位置とそのパワーの大きさを基にアクセントについての評価を行う処理の少なくとも1つの処理を行うようにしている。

【0009】また、前記話者の発話についての評価を行う際、手本の発話内容を幾つかに区切って複数の区間を設定し、話者の発話内容に対しても、前記パターンマッ

チングによるフレーム間の対応付けを用いて、前記手本の区切りに対応する区切りごとに区切って、それにより設定された区間ごとに、前記発話の適正さを判断するに必要な要素についての評価を行うとともに、全区間に渡って前記発話の適正さを判断するに必要な要素についての評価を行い、さらに、その発話内容全体に対する総合的な評価を行うようしている。

【0010】また、前記話者の発話評価を行う際、話者の音声と手本の音声に対して話者正規化処理を施したのち発話評価処理を行うようにしてもよい。

【0011】また、請求項6に記載された本発明の語学練習用発話評価装置は、手本の発話内容に従って発話した話者の音声を分析して得られた音声データと前記手本の発話内容に対する音声データとを比較して、話者の発話評価を行う語学練習用発話評価装置において、手本の持っているそれぞれの発話内容についてそれぞれのフレームごとに求められた発話の適正さを判断するに必要な要素に対するデータを、それぞれの発話内容ごとに記憶する手本音声データ記憶部と、話者の音声を入力する音声入力部と、この音声入力部に入力された話者の発話した発話内容について、それぞれのフレームごとに音声分析し、前記発話の適正さを判断するに必要な要素に対するデータをフレームごとに求める入力音声分析部と、この入力音声分析部により求められた話者のデータと前記手本のデータとを、パターンマッチングによるそれぞれのフレーム間の対応付けを行ってデータ間を比較し、その比較結果に基づいて話者の発話についての評価を行う発話評価処理部と、この発話評価処理部による評価内容に基づいた評価結果を出力する評価結果出力部とを有する構成としたことを特徴としている。

【0012】また、前記発話の適正さを判断するに必要な要素は、発音、アクセント、抑揚、発話のテンポであって、これらの各要素のうち少なくとも1つを用い、前記発話評価処理部が行う発音についての評価は、前記データとして話者の音声の特徴を表す音声特徴データと手本の音声の特徴を表す音声特徴データを用い、両者の音声特徴データをパターンマッチングによるフレーム間の対応付けを行って比較し、その比較結果を基に話者の発音についての評価を行い、アクセントについての評価は、前記データとして話者のパワー情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のパワー情報を対応づけて比較し、その比較結果を基にアクセントについての評価を行い、抑揚についての評価は、前記データとして話者のピッチ情報を用い、前記音声特徴データの対応付けに用いられた各フレーム間の対応関係を用いて、話者と手本のピッチ情報を対応づけて比較し、その比較結果を基に抑揚についての評価を行い、テンポについての評価は、手本の発話内容を幾つかに区切ることによって設定された複数の区

間ごとの時間情報と、前記パターンマッチングによるフレーム間の対応付けにより前記手本の区切りに対応して区切られて設定された話者の発話内容に対するそれぞれの区間ごとの時間情報を比較し、その比較結果をもとに発話のテンポの評価を行うようにしている。

【0013】そして、前記発話評価処理部は、アクセントについての評価を行う際、話者のパワーと手本のパワーとをそれぞれのフレームごとに差を取り、その差分を基にアクセントについての評価を行う処理と、話者のパワーの極大値と手本のパワーの極大値とを比較し、その時間的位置とそのパワーの大きさを基にアクセントについての評価を行う処理の少なくとも1つの処理を行うようしている。

【0014】また、前記発話評価処理部は、手本の発話内容を幾つかに区切って複数の区間を設定し、話者の発話内容に対しても、前記パターンマッチングによるフレーム間の対応付けを用いて、前記手本の区切りに対応する区切りごとに区切って、それにより設定された区間ごとに、前記発話の適正さを判断するに必要な要素についての評価を行うとともに、全区間に渡って前記発話の適正さを判断するに必要な要素についての評価を行い、さらに、その発話内容全体に対する総合的な評価を行うようしている。

【0015】さらに、前記発話評価処理部の前段に、話者の音声と手本の音声に対して話者正規化処理を施す話者正規化処理手段を設け、この話者正規化処理手段により話者正規化処理を行ったのち、発話評価処理を行うようとしてもよい。

【0016】また、請求項11に記載された本発明の発話評価処理プログラムを記憶した記憶媒体は、手本の発話内容に従って発話した話者の音声を分析して得られたデータと前記手本の発話内容に対するデータとを比較して、話者の発話評価を行う発話評価処理プログラムを記憶した記憶媒体において、前記発話評価処理プログラムは、手本の持っているそれぞれの発話内容ごとに、その発話内容についてそれぞれのフレームごとに音声分析し、発話の適正さを判断するに必要な要素に対するデータを予め求めて記憶させておき、話者の発話した発話内容についてそれぞれのフレームごとに音声分析し、前記発話の適正さを判断するに必要な要素に対するデータを求める、この話者のデータと前記手本のデータとを、パターンマッチングによる各フレーム間の対応付けを行ってデータ間を比較し、その比較結果に基づいて話者の発話評価を行うことを特徴としている。

【0017】本発明は、英語など他国言語での会話を効率よく練習可能とするものである。これを実現するために、本発明は、手本の持っているそれぞれの発話内容ごとに、発話の適正さを判断するに必要な要素に対するそれぞれのデータを予め求めて記憶させておき、話者の発話した発話内容について、前記発話の適正さを判断する

に必要な要素に対するデータを求め、両者のデータを、パターンマッチングによる各フレーム間の対応付けを行って比較することで、話者の発話評価を行うようにしている。これにより、話者の発話に対して客観的な評価が可能となり、しかも、発話の適正さを判断するに必要な複数の要素それぞれについて評価するので、正確な評価を行うことができる。

【0018】なお、この発明では、発話の適正さを判断するに必要な複数の要素として、発音、アクセント、抑揚、さらに、発話のテンポを用いている。したがって、各国の言語に特徴的な要素についての評価を行うことができる。

【0019】また、その評価を行う際、ある1つの発話内容において、各要素ごとの評価、その発話内容の各区間ごとの評価、その発話内容全区間の評価、さらには、これらを総合した総合評価も可能となり、ユーザは、自分の発話について、よい部分と悪い部分をきめ細かに知ることができ、効率のよい会話練習を行うことができる。

【0020】さらに、アクセントの評価を行う際、話者のパワーと手本のパワーとをそれぞれのフレームごとに差を取り、その差分を基にアクセントについての評価を行う処理と、話者のパワーの極大値と手本のパワーの極大値とを比較し、その時間的位置とそのパワーの大きさを基にアクセントについての評価を行う処理を組み合わせた処理を行うようにすることで、より一層、適正なアクセントについての評価が可能となり、特に、話者のパワーの極大値と手本のパワーの極大値とを比較することで、1つの発話内容全体のアクセントだけではなく、部分的なアクセントについて評価することもでき、重要な部分のアクセントについての練習を効率よく行うことができる。

【0021】加えて、前記話者の発話評価を行う際、話者の音声と手本の音声とを話者正規化処理を行ったのちに発話評価処理を行うことにより、得られる評価値の信頼性をより一層向上させることができる。

【0022】

【発明の実施の形態】以下、本発明の実施の形態を図面を参照して説明する。なお、この実施の形態では、他国言語として英語を例にとり、英会話の練習を行う例について説明する。

【0023】図1は本発明の実施の形態による英会話練習評価の概略的な処理手順を説明するフローチャートであり、以下、このフローチャートを参照しながら概略的な処理手順について説明する。

【0024】図1において、まず、音声入力の有無を判定し（ステップs1）、音声入力が有れば、音声区間を切り出し、その音声区間を分析する（ステップs2）。この音声分析処理は、ユーザの発話する会話内容について音声特徴量、音量（以下、パワーという）、ピッチな

どの情報、さらには、発話内容を幾つかの区間に区切った区間ごとの時間情報などを検出する処理であり、これら各情報については後に詳細に説明する。

【0025】これら音声特徴量、パワー、ピッチ、時間情報のうち、まず、音声特徴量を用いて、ユーザの音声特徴量と手本の音声特徴量とを、パターンマッチングによる対応付けを行って発音についての評価を行う(ステップs3)。

【0026】そして次に、ユーザの音声波形から求められたパワーの情報と手本のパワー情報を、前述した音声特徴量の対応付けを行ったときのパターンマッチングの対応付けを用いて比較し、アクセントについての評価を行う(ステップs4)。さらに、ユーザの音声波形から求められたピッチに関する情報と手本のピッチ情報を、前述した音声特徴量の対応付けを行ったときのパターンマッチングの対応付けを用いて比較し、抑揚についての評価を行う(ステップs5)。

【0027】次に、発話のテンポについての評価を行う(ステップs6)。この発話のテンポの評価は次のようにして行う。まず、発話内容の手本の発話内容を予め幾つかの区間に区切っておき、音声特徴量を求めたときのパターンマッチングの対応付けを用いて手本の発話内容の区切り点に対応するユーザの発話した発話内容の区切り点を求め、手本の区切り点間の時間的長さと、それに対応する区間におけるユーザの区切り点間の時間的長さを比較して、発話のテンポを評価する。

【0028】そして、ステップs3～s6で求められた4つの評価をもとに話者の発話評価処理を行い(ステップs7)、その評価結果を数字やグラフ等で表示する(ステップs8)。

【0029】以上の処理内容はこの実施の形態を概略的に説明したもので、以下に、詳細に説明する。

【0030】ステップs2における音声分析処理は、発話の適正さを判断するに必要な要素に対して評価値を求めるための各種データを抽出するもので、その要素としては、発音、アクセント、抑揚、テンポであり、これらの要素を評価するに必要なデータとして、音声特徴量、パワー、ピッチ、各区間ごとの時間情報を抽出する。

【0031】一方、手本の持っているそれぞれの発話内容についても音声分析処理を行っておき、それぞれの発話内容ごとに、音声特徴量、パワー、ピッチ、各区間ごとの時間情報を抽出しておく。

【0032】今、手本の発話内容が「Are you ready？」であったとし、その音声波形が、図2(a)のような波形であったとする。

【0033】このような音声波形をたとえば20 msec(シフト量はたとえば10 msec)というような短時間ごとに区切って、それぞれ区切られた区間を1フレームとし、音声の特徴を表す情報としての音声特徴ベクトル(LPCケプストラム係数など)を、それぞれのフレー

ムごとに求める。また、音声のパワーをそれぞれのフレームごとに求めるとともに、音声のピッチをそれぞれのフレームごとに求める。なお、ピッチは、ここでは、それぞれのフレームごとに平均の周波数を求め、その求められた平均の周波数をそれぞれのフレームにおけるピッチと呼ぶ。

【0034】このようにして求められた手本の発話内容における各フレームごとに求められたパワーを結んだ曲線を図2(b)に示し、手本の発話内容における各フレームごとに求められたピッチを結んだ曲線を図2(c)に示す。なお、ピッチはすべてのフレームごとに求める必要はなく、予め設定された区間のみを比較するようにしてもよい。ここで予め設定された区間というのは、発話内容において特徴的な部分、たとえば、発話内容が疑問文であれば、尻上がりとなる最後のフレーズのみのピッチについてユーザの発話内容と手本の発話内容とを比較する。なお、この具体的な説明は後述する。

【0035】さらに、手本の発話内容を幾つかの区間に区切って、各区間ごとの時間情報を得る。この発話内容を幾つかの区間に区切って、それぞれの区間を求める方法としては、音声波形を見ながらパワーの大きさなどを基にして、1つの発話内容を幾つかの区間に予め区切つておく。たとえば、手本の発話内容として「Are you ready？」について考えると、図2(a)に示されるように、「are」、「you」、「ready」を3箇所の区切り点p1, p2, p3で区切り、3つの区間(区間1、区間2、区間3)を設定することができる。そして、各区間ごとの時間情報を得る。ここでは、図2(d)に示すように区間1の時間をT1、区間2の時間をT2、区間3の時間をT3とする。

【0036】以上のようにして手本の音声波形に対する音声特徴ベクトル、パワー、ピッチ、各区間ごとの時間情報を、それぞれの会話練習用の発話内容ごとに求めておく。

【0037】一方、「Are you ready？」をユーザが発話した場合の音声波形が図2(e)であったとする。このような音声波形を、前記手本の音声波形と同じく、短時間(20 msec、シフト量は10 msec)ごとに区切って、区切られた区間を1フレームとし、音声の特徴を表す情報としての音声特徴ベクトル(LPCケプストラム係数)を、それぞれのフレームごとに求める。同様に、それぞれのフレームごとのパワーを求めるとともに、それぞれのフレームごとのピッチを求める。さらに、各区間ごとの時間情報を求める。この各区間ごとの時間情報を求める方法については後述する。

【0038】このようにして求められたユーザの発話内容(「Are you ready?」)における各フレームごとに求められたパワーを結んだ曲線を図2(f)に示し、各フレームごとに求められたピッチを結んだ曲線を図2(g)に示す。

【0039】次に、ユーザの発話内容をもとに求められた各フレームごとのパラメータのうち、まず、音声特徴ベクトル（LPCケプストラム係数）を用い、このユーザの音声波形から得られた各フレームごとの音声特徴ベクトルと手本の音声波形から得られた各フレームごとの音声特徴ベクトルとを、DPマッチングを行ってそれぞれの音声特徴ベクトル間の対応付けを行い、両者の距離を求める。このDPマッチングによる対応付けを行う際、各フレームごとのLPCケプストラム係数だけでなく、 Δ （デルタ）LPCケプストラム係数も用いる。

【0040】この Δ LPCケプストラム係数は、ここでは、隣接するフレームのLPCケプストラム係数との差分のLPCケプストラム係数とするが、その求め方はこれに限られるものではない。このように、パターンマッチング（ここではDPマッチング）による対応付けを行う際、LPCケプストラム係数だけではなく Δ LPCケプストラム係数をも用いることにより、より一層、高精度な対応付けが可能となる。

【0041】そして、以上の処理により求められた手本の発話データとユーザの発話データとの距離を基にして発音の評価値を求める。なお、この評価値を出す際、発話内容全体の評価値だけでなく、前述した区間1、区間2、区間3それぞれにおける評価値を出すようにする。このそれぞれの区間を求める場合、手本の発話内容については、前述したように、音声波形を見ながらパワーの大きさなどを基にして、手本の1つの発話内容を幾つかの区間に予め区切るが、ユーザの発話内容については、この手本の区切り点p1, p2, p3に基づいて自動的に区切ることができる。

【0042】たとえば、ユーザが「Are you ready？」と発話したとき、その発話内容について手本の3箇所の区切り点p1, p2, p3に対応する区切り点p1, p12, p13を抽出する。これは、音声特徴ベクトルの対応付けを行うために用いたフレーム間の対応関係を用いることによって、手本の3箇所の区切り点p1, p2, p3に対応する区切り点p11, p12, p13を求めることができる。そして、これらの区切り点p11, p12, p13によりユーザの発話内容の区間1、区間2、区間3が得られ、図2(h)に示すように、区間1における時間T11、区間2における時間T12、区間3における時間T13を得ることができる。

【0043】ところで、これら区間1、区間2、区間3において、前述したようなDPマッチングによる音声特徴ベクトルの各フレームごとの対応付けによる評価値（発音についての評価値）をそれぞれ求める。一例として、図3に示すように、区間1では評価値0.8、区間2では評価値0.9、区間3では評価値0.5というようにそれぞれの区間ごとの評価値が求められたとする。また、その発話内容全体（区間1から区間3まで）の評価値として0.7という値が求められたとする。

【0044】以上の音声特徴ベクトル（LPCケプストラム係数）を用いてDPマッチングを行い、ユーザと手本の音声データ間の距離を求めてその距離に基づく評価値を算出する処理は、図1におけるステップs3の処理であり、このステップs3の処理により、ユーザの発音が適正か否かの評価を行うことができる。なお、この実施の形態では、評価値が高いほど適正な発音であるものとする。

【0045】次に、パワーを用いてアクセントについての評価を行う。このパワーを用いてのアクセントの評価は、図1におけるステップ4の処理であり、以下のようにして行う。

【0046】まず、前述した音声特徴ベクトルの対応付けを行うために用いたフレーム間の対応関係を用いて、ユーザの発話内容の各フレームのパワーと手本の発話内容の各フレームのパワーとを付けし、両者の差分をそれぞれのフレームごとに求める。そして、各フレームごとにパワーの差分が求められると、求められたそれぞれの差分の和（たとえば、差分の2乗の和）を求め、この差分の和の値に基づく評価値を求めて、その評価値により、ユーザの発話内容のアクセントが適正か否かの評価を行う。また、このとき、ステップs3と同様、 Δ 成分（この場合は、 Δ パワー成分）を用いれば、より一層、高精度な評価を行うことができる。

【0047】なお、このとき、ステップs3と同様に、区間1、区間2、区間3のそれぞれの区間についても、対応するフレーム間のパワーの差分を取り、その区間内における差分の和を求めて、求められた差分の和の大きさから区間ごとの評価値を求める。

【0048】また、パワーによる評価は以下に示すような方法でにおいても評価を行うこともできる。

【0049】すなわち、パワーは図2(b), (f)に示す曲線からもわかるように、時間方向に対してその大きさが変化するが、パワーの極大値が得られる時刻（時間軸上における位置）に重要性の高い部分的なアクセントが有ると判断して、パワーの極大値の時刻とそのパワーの大きさを用いて、重要性の高い部分的なアクセントの位置の評価を行うこともできる。この図2(b)の例における2つの極大値m1, m2と、ユーザの発話内容に対する極大値m11, m12(図2(f))を比較したとき、それぞれの極大値の時間的位置とその大きさが比較的近いと判断されれば、高い評価値となる。

【0050】したがって、前述のパワーの差分の和から求めた評価値による評価と、パワーの極大値とそのパワーの大きさから求めた部分的なアクセントの位置の評価を組み合わせれば、ある発話内容についてその発話内容が適正なアクセントで発話されているか否かを適正に評価できる。

【0051】次に、ピッチによる抑揚の評価について説明する。これは、図1のステップs5における処理であ

る。

【0052】まず、前述のステップs 3において音声特徴ベクトルの対応付けを行うために用いたフレーム間の対応関係を用いて、ユーザの発話内容の各フレームのピッチと手本の発話内容の各フレームのピッチとを対応付けし、両者を比較してその比較結果に基づいた評価値を求める。

【0053】なお、このピッチに関しては、前述したように、すべてのフレームについて行う必要は必ずしもなく、予め設定された区間のみを比較するようにしてもよい。

【0054】ここでの予め設定された区間というのは、発話内容において特徴的な部分、たとえば、発話内容が疑問文で有れば、尻上がりとなる最後のフレーズのみのピッチについてユーザの発話内容と手本の発話内容とを比較する。

【0055】図2における「Are you ready ?」の発話内容を例に取れば、区間3の「ready？」の部分について両者のピッチを比較する。

【0056】手本の発話内容から求められた区間3のピッチ(図2(c))とユーザの発話内容から求められた区間3のピッチ(図2(g))を比較すると、図2(c)は尻上がりな変化となっているのに対し、図2(g)は尻下がりな変化となっている。したがって、ユーザの発話内容は、手本の発話内容に対して、「ready？」の部分の抑揚が大きく異なると判断でき、これにより、ピッチに関する区間3の評価値は低い値となる。

【0057】次にユーザの発話した発話内容全体のテンポが手本の発話内容に対して適正か否かを評価する処理について説明する。これは、図1のステップs 6における処理である。

【0058】この発話内容全体のテンポが適正か否かを判断する処理は、以下のようにして行う。

【0059】手本の3箇所の区切り点p1, p2, p3に対応するユーザの区切り点p11, p12, p13を求めたとき、ユーザの区切り点p11, p12, p13と、手本の区切り点p1, p2, p3との時間的位置を基にして、区間1、区間2、区間3の時間を比較することにより、区間ごとのテンポが適正か否か、あるいは、その発話内容全体(区間1～3)のテンポが適正か否かを評価する。

【0060】具体的には、たとえば、手本の発話内容の区切り点p1に対して、ユーザの発話内容の区切り点p11の時間的位置が、時間的に早い位置、つまり、ユーザの区間1の時間T11が手本の区間1の時間T1よりも少し短いとすると、区間1におけるテンポが手本に比べて少し速すぎるということが言える。したがって、この場合の、区間1におけるテンポに関する評価値は少し低い値となる。

【0061】この図2の例では、同図(d)と(h)を

比べると、ユーザの区間3の時間T13が手本の区間3の時間T3よりも長く、区間3におけるテンポが手本に比べて遅すぎることが言える。同様にして、各区間についても評価を行い、それぞれの区間における評価値を求め、さらに、発話内容全体についての評価値を求める。

【0062】また、テンポに関する評価については、前述のように求められた、手本の音声の各区間の時間長の比T1:T2:T3と、ユーザーの発話した音声の各区間の時間長の比T11:T12:T13とを比較することにより、評価値を求めるこどもできる。

【0063】以上のようにして、ユーザの発話した発話内容における発音に関する評価として、区間1～3についての評価値と発話内容全体についての評価値とを求める(ステップs 3)、アクセントに関する評価として、区間1～3についての評価値と発話内容全体についての評価値とを求める(ステップs 4)、抑揚に関する評価として、区間1～3についての評価値と発話内容全体についての評価値とを求める(ステップs 5)。なお、この例では、抑揚については区間3のみについて評価を行っている。

【0064】そして、これらの処理により得られた評価値を基に、総合評価を行う(ステップs 6)。たとえば、発音に関しては、図3に示すように、区間1における評価値が0.8、区間2における評価値が0.9、区間3における評価値が0.5、発話内容全体(区間1～3)における評価値が0.7であったとする。また、アクセントに関しては、区間1における類似度が0.7、区間2における評価値が0.9、区間3における評価値が0.7、発話内容全体における評価値が0.8であったとする。また、抑揚に関しては、区間3のみについて判断しているので、区間3における評価値が0.3、発話内容全体における評価値も0.3であったとする。また、テンポに関しては、区間1における評価値が0.6、区間2における評価値が0.8、区間3における評価値が0.4、発話内容全体における評価値が0.6であったとする。

【0065】そして、これらの値をもとにユーザの発話内容についての総合評価として、たとえば、0.5というような総合評価値(たとえば、0.1～1.0まで0.1刻みの評価値とする)で出される。

【0066】このような発音、アクセント、抑揚、テンポなどについて、各区間ごとの評価値、発話内容全体(区間1～3)の評価値、これらすべてを総合した総合評価値は、ユーザに対して表示されるようになっていく。

【0067】ユーザは、その表示結果を見ることによって、「Are you ready ?」という発話内容全体の総合評価値は「0.5」と比較的低い値であることがわかつ

り、また、発音、アクセント、抑揚、テンポについて、どの区間がよくてどの区間が悪いかなど、発話内容についてきめ細かな評価を知ることができる。

【0068】そして、ユーザは、このような評価結果を踏まえて悪いところを注意しながら、同じ発話内容（この場合、「Are you ready？」）について、手本を聴きながらなるべく手本と同じようになるように繰り返し発話すると、段々と総合評価値が上昇して行き、たとえば、2回目の総合評価値が0.7、3回目が0.9という具合に総合評価値が段々と上昇して行く。そして、さらに、数回繰り返して発話したとき、総合評価値が或る高い値（たとえば、0.9）で一定となったとき、その発話内容に関しては、発音、アクセント、抑揚、テンポなどを総合して良好な発話がなされていると判断できる。このように、1つの発話内容について何回も練習した結果、総合評価値が或る高い値に収束するようになると、良好な発話が行えるようになったと判断する。

【0069】なお、前記発音、アクセント、抑揚、テンポの4つの要素についてのそれぞれの評価値に重み付けを行ってもよい。たとえば、ある発話内容に対して、発音についての発話内容全区間（区間1～区間3）の評価値にw1という重みを掛け算し、アクセントについての発話内容全区間（区間1～区間3）の評価値にw2という重みを掛け算し、抑揚についての発話内容全区間（区間1～区間3）の評価値にw3という重みを掛け算し、テンポについての発話内容全区間（区間1～区間3）の評価値にw4という重みを掛け算して、それぞれ重み付けをした評価値を算出したのち、総合評価値を出すようにしてもよい。これら、重みw1, w2, w3, w4は、発音、アクセント、抑揚、テンポの4つの要素についてそれぞれの重要度などにより予め決定しておく。このような重み付けを行うことにより、より一層、適正な評価を行うことができる。さらに、この重み付けは、発音、アクセント、抑揚、テンポの4つの要素について各区間ごとに行つてもよい。

【0070】ところで、手本の発話内容に対して、ユーザが異なる発話内容を入力してしまった場合も、場合によっては、総合評価値として比較的高い評価値が出てきてしまう可能性もある。このような場合、前述した区切り点の時間的位置を基に手本の発話内容内とは異なる発話内容であることを判定する。

【0071】たとえば、手本の発話内容「Are you ready？」に対してユーザが異なった発話をした場合、図4に示すように、両者の音声波形は大きく異なるものとなり、たとえ、全体的なフレーム数にあまり大きな差がなくても（例として、手本のフレーム数が70個、ユーザの発話内容のフレーム数が80個とする）、手本の区切り点p1, p2, p3に対応する区切り点をユーザの発話内容に対応させて無理矢理区切ると（区切り点p11, p12, p13）、それぞれの区間1、区間

2、区間3の時間的な長さ（フレーム数）に大きな違いの出る場合が多い。この図4の例では、区間1におけるユーザの発話時間が手本の約2倍となっている。

【0072】このように、手本の発話内容ごとに求められている幾つかの区間にに対して、ユーザの発話した発話内容に時間的長さの大きく異なった区間（たとえば、2倍以上の長さを有する場合など）が存在する場合は、ユーザが手本の発話内容とは異なった発話内容を入力した可能性が高いということを判定できる。

10 【0073】なお、以上説明した処理によっても、ユーザの発話する発話内容に対して信頼性の高い評価値を得ることが可能であるが、話者正規化処理を行ったのち、前述したそれぞれの評価値を求める処理を行うようにしてもよい。

【0074】たとえば、予め話者適応用の数単語をユーザに発話してもらい、それにより得られた音声データと手本が持っているその単語についての音声データとを基にして変換関数を作成し、その変換関数を用いてユーザの音声を手本の音声側（あるいはその逆）にシフトして話者正規化処理を行う。このような話者正規化を行ったのちに、前述したような評価値を求める処理を行えば、より一層、高精度な評価が可能となる。

【0075】図5は以上説明した発話内容評価処理を行うための装置構成例を示すもので、音声入力部1、入力音声分析部2、手本音声データ記憶部3、変換関数作成部4、変換関数記憶部5、話者正規化処理部6、発話評価処理部7、評価結果出力部8などから構成されている。

30 【0076】音声入力部1はマイクロホンやA/D変換器などから構成され、ユーザの発話した音声を入力してA/D変換する。また、入力音声分析部2は、ユーザの発話した入力音声を分析して、音声特徴量としてのLPCケプストラム係数や、音声レベル、ピッチを各フレームごとに求め、さらには、幾つかの区間を設定してそれぞれ区間ごとの時間情報を求める。

【0077】また、手本音声データ記憶部3は、手本となる複数の発話内容それぞれについて、それぞれのフレームごとのLPCケプストラム、パワー、ピッチ、各区間ごとの時間情報などを記憶するものである。

40 【0078】また、変換関数作成部4は、ユーザの音声特徴量と手本の音声特徴量を用いて変換関数を作成するもので、これは前述したように、予め用意された数単語を用いて予め変換関数を作成しておく。そして、その変換関数を変換関数記憶部5に記憶しておく。

【0079】話者正規化処理部6は、変換関数記憶部5に記憶されている変換関数を用いてユーザの音声の特徴を手本の音声の特徴に近づける処理を行う。この話者正規化処理を行うことで、ユーザの音声の話者性と手本の音声の話者性との差を小さくすることができるため、その後に行う発話評価処理をより一層高精度に行なうことが

できる。

【0080】発話評価処理部7は、ユーザの音声データから得られた各フレームごとの音声特徴ベクトルと手本の音声波形から得られた各フレームごとの音声特徴ベクトルとを、DPマッチングを行ってそれぞれの特徴ベクトル間の対応付けを行い、両者の距離を求め、その距離を基に、発音の適正さを表す指標としての評価値を求めるとともに、音声レベル情報を用いてアクセントについての評価を行うための評価値を求め、さらに、ピッチ情報を用いて抑揚の評価を行うための評価値を求め、さらに、ユーザの発話した発話内容全体のテンポが手本の発話内容に対して適正か否かを評価するための評価値を求める処理を行う。

【0081】そして、この発話評価処理部7で求められた発音、アクセント、抑揚、テンポ、などに関するそれぞれの評価値が求められると、評価結果出力部8により求められた評価値を表示する。

【0082】なお、発話評価処理部7による評価値算出処理とその具体的な評価値の例については既に説明したので、ここではその詳細な説明は省略する。

【0083】なお、この図4の構成は話者適応を行わせるために、変換関数作成部4、変換関数記憶部5、話者適応化処理部6を設けた構成としたが、これは必ずしも必要ではなく、これらを設けなくても本発明の所期の目的は達成できるものである。

【0084】また、本発明は以上説明した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、前述の実施の形態では、発話の適正さを判断するに必要な複数の要素として、発音、アクセント、抑揚、テンポの4種類すべてを用いた例について説明したが、必ずしもこれらすべてを用いなくてもよい。たとえば、発音、アクセントのみを評価するということもでき、これにテンポを加えるなど、それぞれの場合によって各要素を組み合わせて用いることによって、より一層、適正な発話評価が可能となることは勿論である。

【0085】また、発音の評価を行うための音声の特徴ベクトルの対応付けをDPマッチングを用いて行う例について説明したが、これは、DPマッチングに限られるものではなく、隠れマルコフモデル(HMM)や、ニューラルネットワークによるマッチングを用いることもできる。

【0086】また、以上説明した本発明の処理を行う処理プログラムは、フロッピディスク、光ディスク、ハードディスクなどの記憶媒体に記憶させておくことができ、本発明はその記憶媒体をも含むものである。また、ネットワークから処理プログラムを得るようにしてよい。

【0087】

10

【発明の効果】以上説明したように、本発明によれば、手本の持っているそれぞれの発話内容ごとに、発話の適正さを判断するに必要な複数の要素に対するそれぞれの音声データを予め求めて記憶させておき、話者(ユーザ)の発話した発話内容について、前記複数の要素に対する音声データを求める、両者の音声データを、パターンマッチングによる各フレーム間の対応付けを行って比較することで、話者の発話評価を行うようしている。これにより、話者の発話に対して客観的な評価が可能となり、しかも、発話の適正さを判断するに必要な複数の要素について評価するので、正確な評価を行うことができる。

【0088】また、この発明では、発話の適正さを判断するための指標となる複数の要素として、発音、アクセント、抑揚を用い、さらに、発話のテンポも用いているので、各国の言語に特有な発音、アクセント、抑揚、テンポについてそれぞれ評価を行うことができ、それぞれの国の言語に適応した効率的な会話練習が可能となる。

20

【0089】また、その評価を行う際、ある1つの発話内容において、各要素ごとの評価、その発話内容の各区間ごとの評価、その発話内容全区間の評価、さらには、これらを総合した総合評価も可能となり、話者は、自分の発話について、評価の高い部分と評価の低い部分をきめ細かに知ることができ、効率のよい会話練習を行うことができる。さらに、この発話評価を行う際、話者の音声と手本の音声とを話者正規化処理を行ったのちに発話評価処理を行うことにより、得られる評価値の信頼性をより一層向上させることができる。

30

【0090】さらに、アクセントの評価を行う際、話者のパワーと手本のパワーとをそれぞれのフレームごとに差を取り、その差分を基にアクセントについての評価を行う処理と、話者のパワーの極大値と手本のパワーの極大値とを比較し、その時間的位置とそのパワーの大きさを基にアクセントについての評価を行う処理を組み合わせた処理を行うようにすることで、より一層、適正なアクセントについての評価が可能となり、特に、話者のパワーの極大値と手本のパワーの極大値とを比較することで、1つの発話内容全体のアクセントだけではなく、部分的なアクセントについて評価することもでき、重要な部分のアクセントについての練習を効率よく行うことができる。

40

【0091】加えて、前記話者の発話評価を行う際、話者の音声と手本の音声とを話者正規化処理を行ったのちに発話評価処理を行うことにより、得られる評価値の信頼性をより一層向上させることができる。

【図面の簡単な説明】

【図1】本発明の実施の形態の発話評価処理手順を概略的に説明するフローチャート。

【図2】本発明の実施の形態の発話評価処理を説明する

50

タイムチャート。

【図3】本発明の実施の形態において求められた評価値の一例を示す図。

【図4】本発明の実施の形態において、ユーザが手本と異なる発話内容を発話した場合の対処方法を説明する図。

【図5】本発明の語学練習用発話評価装置の実施の形態を示すブロック図。

【符号の説明】

- 1 音声入力部
- 2 入力音声分析部
- 3 手本音声データ記憶部

4 変換関数記憶部

5 変換関数記憶部

6 話者正規化処理部

7 発話評価処理部

8 評価結果出力部

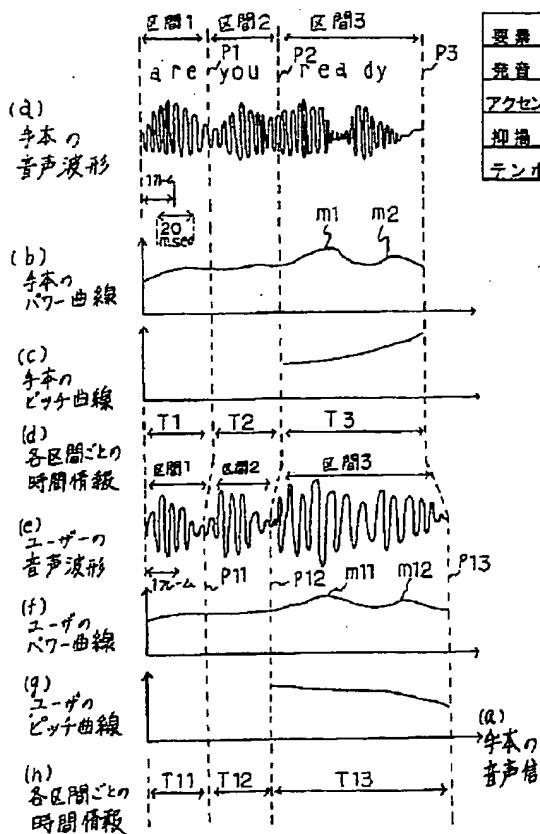
p1, p2, p3 手本の発話内容の区切り点

p11, p12, p13 ユーザの発話内容の区切り点

T1, T2, T3 手本の発話内容における各区間の時間情報

10 T11, T12, T13 ユーザの発話内容における各区間の時間情報

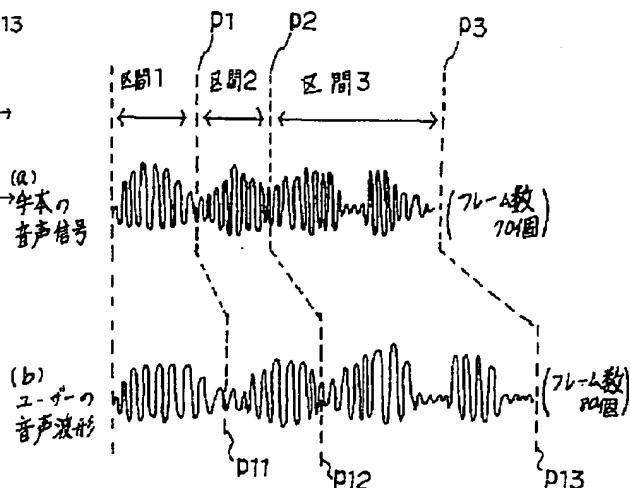
【図2】



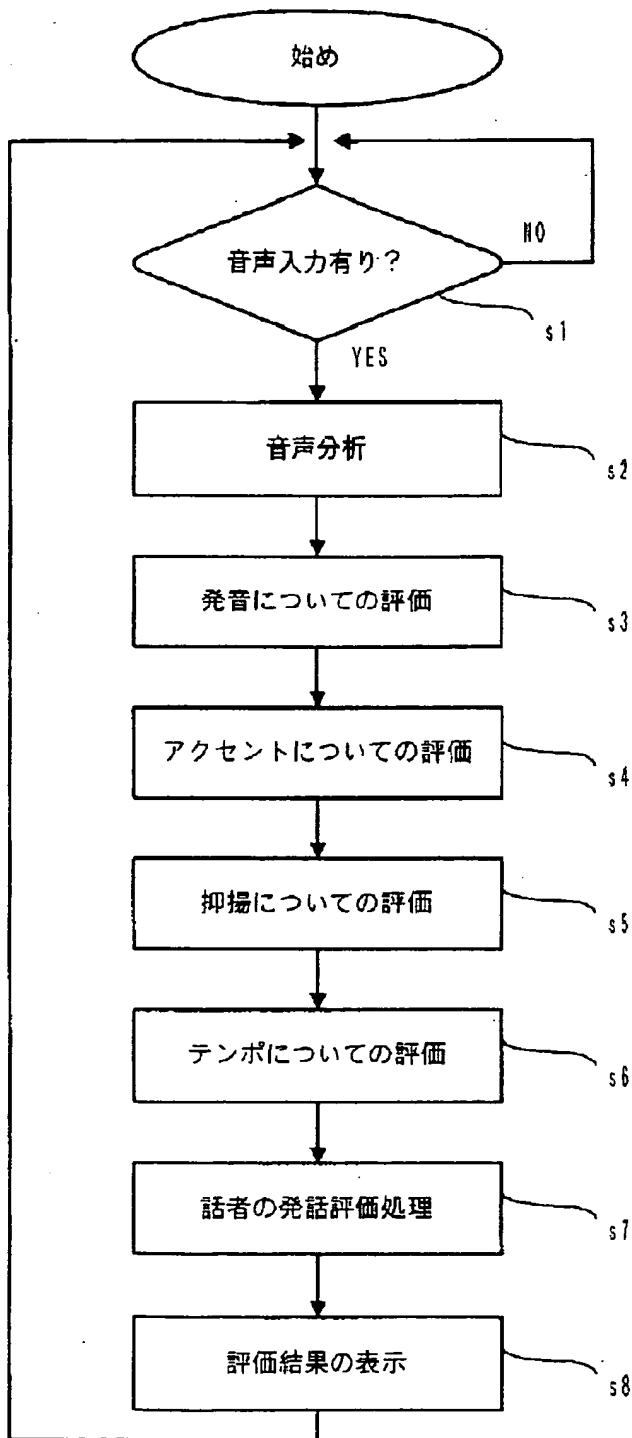
【図3】

要素	区間1の評価値	区間2の評価値	区間3の評価値	区間1~3の評価値	総合評価値
発音	0.8	0.9	0.5	0.7	0.5
アクセント	0.7	0.9	0.7	0.8	
抑揚			0.3	0.3	
テンポ	0.8	0.8	0.4	0.6	

【図4】



[図1]



[図5]

